面向无监督分割的双分支上采样域自适应网络 *

韩宗桓 [1]，刘名果 [1†]，李 珅 [2]，陈立家 [1]，田 敏 [1]，兰天翔 [1]，梁 倩 [1]

(1. 河南大学 物理与电子学院，河南 开封 475004; 2. 开封平煤新型炭材料科技有限公司，河南 开封 475004)

摘 要： 工业应用中，表面压印字符图像全监督语义分割将会给企业带来高昂的数据集标注成本。针对该问题，提出了双支路特征融合的域适应分割方法(Dual-branch Feature Fusion Domain Adaptation，DbFFDA)。首先，借鉴 U-Net 的跨层连接设计思路，提出了双分支上采样结构的残差域适应分割网络(Residual Adaptation Network，Res-Adp)。同时，提出了融合特征输入用于提升网络分割性能，克服了字符缺失的问题。此外，提出了分割连续性损失函数 L_{Con}，抑制了分割图像中噪点的产生。在石墨电极表面压印字符无监督分割实验中，所提方法 MIoU 值可达 69.60%。实际分割效果已基本满足字符识别需求，有望在特定工业场景中投入实际应用，为企业节省巨大的数据集标注成本。

关键词： 表面压印字符；域适应；语义分割；无监督训练

中图分类号： TP391.41 doi: 10.19734/j.issn.1001-3695.2022.02.0062

Double branch upsampling domain adaptive network for unsupervised segmentation

Han Zonghuan[1], Liu Mingguo[1†], Li Shen[2], Chen Lijia[1], Tian Min[1], Lan Tianxiang[1], Liang Qian[1]

(1. *School of Physics & Electronics*, *Henan University*, *Kaifeng Henan* 475004, *China*; 2. *Kaifeng Pingmei New Carbon Materials Technology Co*, *Ltd*, *Kaifeng Henan* 475004, *China*)

Abstract: In industrial applications, fully supervised semantic segmentation of surface-imprinted character images will bring high dataset annotation costs to enterprises. To solve this problem, this paper proposes a dual-branch feature fusion domain adaptation (DbFFDA) method. First, the method draws on the design idea of U-Net's cross-layer connection, and proposes a residual adaptation network (Res-Adp) with a dual-branch upsampling structure. At the same time, this method proposes fusion feature input to improve network segmentation performance and overcome the problem of missing characters. Furthermore, this method proposes a segmentation continuity loss function L_{Con}, which suppresses the generation of noise in segmented images. In the unsupervised segmentation experiment of imprinted characters on the surface of graphite electrodes, the MIoU value of the proposed method can reach 69.60%. The actual segmentation effect has basically met the needs of character recognition, and it is expected to be put into practical applications in specific industrial scenarios, saving enterprises huge dataset labeling costs.

Key words: surface imprint characters; domain adaptation; semantic segmentation; unsupervised training

0 引言

图像语义分割是计算机视觉领域重要的研究方向，在自动驾驶、字符识别等领域具有广泛的应用前景，其任务目标是将图像中各像素点归纳解析为具有特定语义信息的类别。在工业应用中，因受光线、打印机质量影响，表面压印字符图像的分割难度较大，传统算法很难对其进行精确的分割。自深度学习技术问世以来，众多研究者在图像语义分割领域进行了一系列探索，分割效果及精度不断提升。2014 年 Long 等人[1]提出的(Fully Convolutional Networks，FCN)网络完全由卷积层构成，不包含全连接结构，为语义分割网络的设计提供了新思路。2015 年提出的 U-Net[2]设计了一个编码与解码部分完全对称的 U 型网络结构，并引入了跨层连接融合网络各层级间的信息，在医学影像分割领域取得了巨大成功。2018 年，Chen 等人[3~6]提出了 DeepLab 系列网络，其最重要的贡献是提出了空洞空间池化金字塔(Atrous Spatial Pooling Pyramid, ASPP)模块。卷积神经网络(Convolutional Neural Networks，CNN)的出现，极大的提升了复杂场景图像语义分割的效果，具有传统语义分割方法无可比拟的优点。

然而，上述各网络均采用全监督的方式进行训练，即需要手工标注的标签作为训练数据，带来了巨大的手工标注成本。随着域自适应无监督语义分割网络的研究逐渐兴起，这一问题有望得到解决。域适应任务的重点是优化对齐两域数据的特征分布，提取两域数据的共有特征对目标域数据进行预测。研究者们提出了各种域适应方法，其中大多数域适应方法均使用生成对抗网络(Generative Adversarial Networks，GAN)[7]对抗学习进行域适应训练。将分割网络作为生成器，源域和目标域图像送入分割网络(生成器)得到的分割预测交替送入判别器中，分割网络与判别器对抗训练以对齐两域特征的分布，实现域适应分割。Ganin 等人提出的(Domain-Adversarial Neural Network，DANN)[8]由特征提取模块、分类模块和域判别模块组成。该网络同时构造了图像分类损失和域分类损失等两个损失函数，在训练过程中既提升了网络性能又对齐了两域数据的分布，提升了对目标域图像的分类能力。因使用单域判别器的对抗性域适应方法无法利用复杂的多模结构，Pei 等人[9]提出的(Multi-Adversarial Domain

收稿日期： 2022-02-11；**修回日期：** 2022-04-11 **基金项目：** 国家自然科学基金资助项目(61901158)；河南省科技厅项目(202102210121，212102210500)；开封市重大科技专项(20ZD014，2001016)

作者简介： 韩宗桓(1996-)，男，河南新乡人，硕士，主要研究方向为语义分割；刘名果(1984-)，男(通信作者)，河南郑州人，副教授，硕导，博士，主要研究方向为深度学习与演化计算(liumingguo@henu.edu.cn)；李珅(1982-)，男，河南开封人，助理工程师，本科，主要研究方向为语义分割；陈立家(1979-)，男，河南开封人，副教授，硕导，博士，主要研究方向为智能计算；田敏(1996-)，女，山东潍坊人，硕士，主要研究方向为深度学习；兰天翔(1994-)，男，河南开封人，硕士，主要研究方向为语义分割；梁倩(1996-)，女，河南洛阳人，硕士，主要研究方向为目标检测.

Adaptation，MADA)构造了多个判别器进行域适应训练。多个域判别器可从各个维度对齐两域数据的分布，提升域适应效果。Luo 等人[10]提出的(Category-Level Adversarial Network, CLAN)构造了类别级别的对抗网络，旨在整体对齐一致性中强化局部语义一致性。Wang 等人[11]提出的(Patch-based Output Space Adversarial Learning, pOSAL)设计了一个轻量且高效的分割网络。并提出一种新的形态感知分割损失来指导网络生成准确和平滑的分割预测。Wang 等人[12]提出了边界和熵驱动的对抗学习(Boundary and Entropy-driven Adversarial Learning，BEAL)域适应框架，以改进模糊边界区域上的分割性能。BEAL 通过鼓励目标域的边界预测与源域相似，以生成更准确的边界。Zhu 等人[13]提出的循环式生成对抗网络(Cycle-Consistent Generative Adversarial Networks，CycleGAN)将两个 GAN 网络级联在一起，并引入了循环一致性损失函数，构造了一个源域图像与目标域图像的双向风格迁移网络，因此可生成接近目标数据分布的模拟样本协助训练。Chen 等人[14]对 CycleGAN 结构进行改进。在目标域到源域的生成器上添加了分割分支，从图像和特征等两个角度对齐源域图像和目标域图像的分布，在医学 MRI 图像和 CT 图像之间的域适应分割上取得了巨大成功。张勋晖等人[15]提出的域适应分割方法在分割网络中引入了空洞空间池化金字塔(ASPP)提取各尺度的图像特征以提升分割性能，并将分割预测的信息熵作用于对抗损失，以减小域偏移。Liu 等人提出的(Source-Free Domain Adaptation，SFDA)[16]网络提供了一个在源域数据集无法公开的应用场景中进行域适应训练的方法。通过从源域模型中恢复和筛选源域数据，仅使用训练完成的源域模型和目标域数据集即可实现域适应。Araslanov 等人[17]提出的轻量化域适应分割方法针对现有网络模型较为复杂、资源消耗大的缺点，摈弃了对抗训练和风格迁移等常用的域适应方法。采用加噪、翻转和缩放等数据增强技术确保跨域的图像语义分割的一致性。Wang 等人提出的(Correlation-Aware Domain Adaptation，CorDA)[18]使用两域通用的自监督深度估计的指导来弥合领域差距。该方法在目标深度估计的帮助下显式学习任务特征相关性以改善预测效果。Saha 等人[19]提出了一种编码视觉任务关系的方法，用于提升无监督域适应网络的性能。提出的跨任务关系层(Cross-Task Relation Layer，CTRL)编码了语义和深度预测之间的任务依赖关系。刘少鹏等人[20]提出了一种两阶段分割网络——CDR-GANs。各分割阶段都包含语义分割网络、生成器和判别器等三部分，训练过程中，判别器指导语义分割网络和生成器学习原图和分割预测的联合概率分布。Li 等人[21]针对现有域适应算法使用共享的源域网络学习跨域的特征表示限制了对未标记的目标域对象的泛化能力的问题。提出了一种可转移语义增强(Transferable Semantic Augmentation，TSA)方法，通过隐式生成关于目标域对象的源域特征来增强网络的适应能力。

本文所提双支路特征融合的域适应分割方法 DbFFDA 在网络结构、图像预处理和损失函数等方面对 U-Net 进行改进，在石墨电极压印字符数据集上取得了较为理想的分割效果，基本满足工业应用需求。DbFFDA 创新之处包含以下 3 个方面：

a) 提出一种双分支上采样结构的无监督语义分割网络 Res-Adp。在 U-Net 的跨层连接中加入了残差模块，以搭建残差支路用于两域特征对齐。网络上采样过程中的各级特征分别通过残差支路与卷积支路进行上采样。残差支路负责特征对齐，以使用域不变特征对图像进行分割，卷积支路负责保留本域特征，以使用本域独有的图像特征对分割细节进行补充。

b) 融合特征输入。针对表面压印字符图像噪点多、字符边缘极为重要的特点，将表面压印字符灰度图像、中值滤波图像和边缘检测图像作为网络输入的 3 个通道，融合送入网络进行训练。

c) 构造分割连续性损失函数 L_{Con} 约束分割网络的训练。根据分割图像中各类对象内部连续的先验知识提出了分割连续性损失函数 L_{Con}，通过约束源域分割图像的生成间接提升目标域分割效果，抑制了字符中的空洞与背景中噪点的产生。

1　双支路特征融合的域适应分割方法 DbFFDA

1.1　域适应分割框架

域适应分割框架如图 1 所示。分割网络为本文所提双分支上采样网络 Res-Adp，为两域数据所共享。Is 为源域(模拟)数据，其标签极易获取，监督信息丰富，具有完备的标签集 Ls。IT 为目标域(真实)数据，无监督信息。

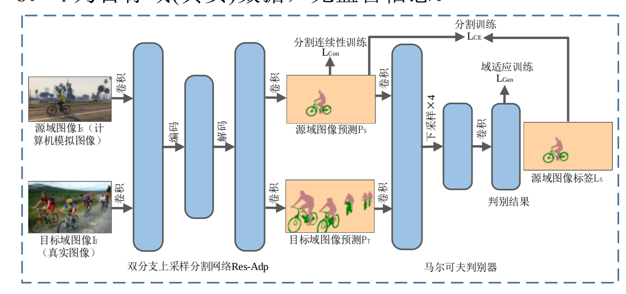


图 1　域适应分割框架

Fig. 1　Domain adaptive segmentation framework

源域数据为计算机生成的模拟图像，标签无须手工标注，目标域数据为摄像机采集到的真实图像，训练数据无标注信息。

a) 源域数据的全监督训练。

(a) 源域(模拟)图像 Is 输入分割网络后得到源域预测 Ps，Ps 可与其标签 Ls 构造交叉熵损失函数，以全监督的方式训练分割网络 Res-Adp。交叉熵损失函数公式为

$$L_{CE}=-\frac{1}{N}\sum_{i=1}^{N}[y^{(i)}\log(p^{(i)})+(1-y^{(i)})\log(1-p^{(i)})] \tag{1}$$

其中，$p^{(i)}$ 为源域预测中的像素点，$y^{(i)}$ 为源域标签中的像素点。

(b) 目标域(真实)图像 IT 输入分割网络后得到目标域预测 PT，因目标域图像 IT 无标签，故不可构造交叉熵损失函数进行分割训练。

b) 分割网络(生成器)与判别器的对抗训练。

为了使得使用源域图像训练的分割网络对目标域图像也能进行精准的分割，需要对齐两域数据的分布，进行域适应训练。将源域预测 Ps 与目标域预测 PT 分别送入判别器，与分割网络(生成器)进行对抗训练。对抗性损失函数 L_{Gan} 如下：

$$L_{Gan}=E_{x_S\sim P_{data}(x_S)}[\log D(S(x_S))]+E_{x_T\sim P_{data}(x_T)}[\log(1-D(S(x_T))] \tag{2}$$

其中，x_S 及 x_T 分别为源域与目标域图像，S 为分割网络，D 为判别器。

网络整体损失函数 L 由三部分构成：

$$L=L_{Gan}+L_{CE}+L_{Con} \tag{3}$$

其中，L_{Gan} 为对抗性损失函数，L_{CE} 为交叉熵损失函数。L_{Con} 为本文所提分割连续性损失函数，将在本文第 2.4 节中详细阐述。

判别器的优化目标是鉴别出输入的分割预测是源域预测 Ps 还是目标域预测 PT。而分割网络(生成器)的优化目标是使得判别器难以甄别分割预测的来源。训练时，判别器训练方向是对源域预测输出全 1 矩阵，对目标域预测输出全 0 矩阵，确定分割预测的类别。判别器对目标域预测的判别结果与同形的全 0 矩阵构造 L2 范数，对源域预测的判别结果与同形

的全 1 矩阵构造 L2 范数，判别器优化减小两损失函数之和。

分割网络优化目标是使得目标域预测送入判别器后输出全 1 矩阵，输出接近源域预测的分割结果。判别器对目标域预测的判别结果与同形的全 1 矩阵构造 L2 范数，源域输入图像与其标签构造非对抗性损失函数，分割网络优化减小两损失函数之和。

非对抗性损失函数 L_{CE} 和 L_{Con} 负责训练分割网络的分割性能，对抗性损失函数 L_{Gan} 负责训练分割网络的域适应能力，在两者的共同约束下，分割网络在提升对源域(模拟)图像分割性能的同时又对齐了两域数据的分布，进而提升了对目标域(真实)图像的分割性能。达到了对目标域(真实)图像无监督分割的目标。

1.2　双分支上采样分割网络 Res-Adp

2015 年提出的 U-Net 设计了一个编码与解码部分完全对称的 U 形网络结构，并引入了跨层连接融合网络各层级间的信息，在医学影像分割领域取得了巨大成功。U-Net 结构如图 2 所示。U-Net 最大的创新点是网络对称位置的跨层连接结构，由于上下采样会不可避免的引起特征丢失，而该网络可以同时利用各个层级的特征进行分割，提升网络性能。残差网络(ResNet)[22]于 2015 年提出，解决了网络深度过深时性能不再提升的问题。网络某层特征图通过两层卷积层提取特征后与该层特图直接相加并激活即得残差块输出。胡扬涛等人提出的 AR-Unet[23]将残差模块引入 U-Net 中，避免了 U-Net 中的梯度消失以及图像结构信息丢失等问题。

U-Net 对于全监督分割任务可以达到较好的分割效果。对于域适应分割任务，源域图像进行全监督训练，分割网络与判别器共同进行对抗训练。U-Net 由于结构过于简单，特征之间直接拼接，无法很好的对齐两域特征。

Res-Adp 网络在 U-Net 编码阶段与解码阶段的跨层连接中引入了残差块对齐两域特征。为了避免因对齐两域特征而丢失本域特征的问题，设计了双分支结构改善该问题。网络上采样过程中的各级特征分别通过残差支路与卷积支路进行上采样。残差支路负责特征对齐，以使用域不变特征对图像进行分割，卷积支路负责保留本域特征，以使用本域独有的图像特征对分割细节进行补充。

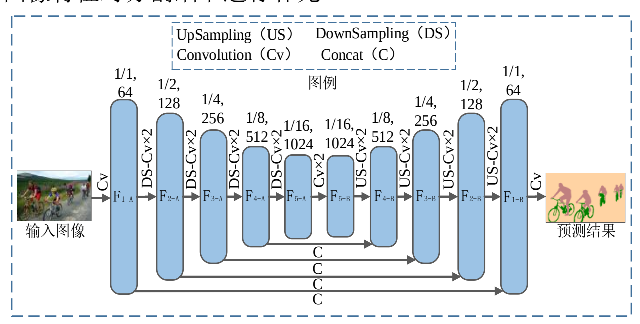


图 2　U-Net 网络结构

Fig. 2　Network structure of U-Net

Res-Adp 结构如图 3 所示，给出了图例，并在连接线上列出了操作的简称。特别的，图中连接线上的 C 表示先执行连接线对应操作再与其他特征图进行拼接。Ri-C 表示先通过 i 个残差块再与其他特征图进行拼接，US-C 表示先上采样再与其他特征图进行拼接。特征图旁的数字，逗号左边的数值表示该特征图边长与网络输入图像边长的比值，逗号右边的数值表示该层特征图的通道数。如 1/2，128 表示该特征图边长为网络输入图像的 1/2，包含 128 个通道。

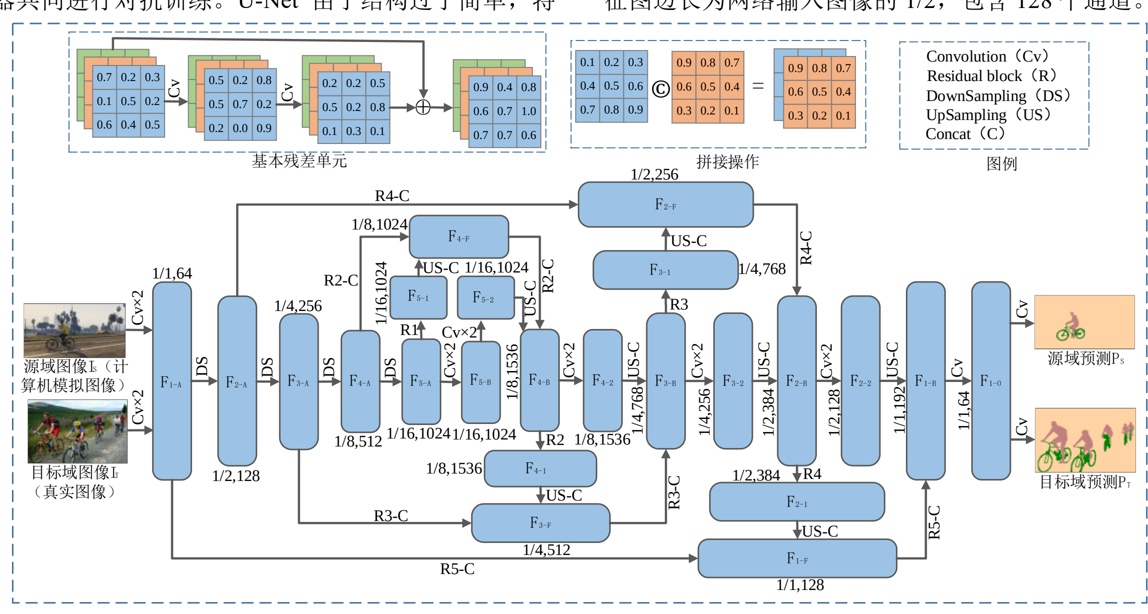


图 3　分割网络 Res-Adp 结构

Fig. 3　Structure of segmentation network Res-Adp

输入图像使用 2 层卷积扩充通道数后进行 4 次下采样到达网络最深层级，按照下采样过程中特征图的尺度将网络划分为 1-5 层，分别记为 $F_{i\text{-}A}$。1 层为输入图像所在层，5 层为网络最深层。

下采样到网络最深层的特征图 $F_{5\text{-}A}$ 卷积 2 次后记为 $F_{5\text{-}B}$。从 $F_{5\text{-}A}$ 和 $F_{5\text{-}B}$ 分别引出一条支路，支路一经 1 个残差块对齐特征后上采样到第 4 层级。与此同时，下采样到第 4 层的特征图 $F_{4\text{-}A}$ 通过 2 个残差块后与支路一拼接，得到融合特征 $F_{4\text{-}F}$。

支路二经 2 层卷积后上采样到第 4 层级，与 $F_{4\text{-}F}$ 通过 2 个残差块后的特征图拼接得到 $F_{4\text{-}B}$。

此后特征图逐步上采样，对于上采样过程中的各层级特征图，均从 $F_{i\text{-}B}$ 引出两条支路。第一支路为残差支路，用于对齐两域特征。第二支路为上采样支路，用于保留本域独有特征。

从 $F_{4\text{-}B}$ 引出两条支路。支路一经 2 个残差块对齐特征后上采样到第 3 层级。与此同时，下采样到第 3 层的特征图 $F_{3\text{-}A}$ 通过 3 个残差块后与支路一拼接，得到融合特征 $F_{3\text{-}F}$。

支路二经过 2 层卷积后上采样到第 3 层级，与 $F_{3\text{-}F}$ 通过 3 个残差块后的特征图拼接得到 $F_{3\text{-}B}$。

同理，重复上述操作，依次得到 $F_{2\text{-}B}$、$F_{1\text{-}B}$。随着上采样过程中网络深度逐步变浅，对齐两域特征所需残差块数量逐步增加，从第 5 层的 1 个残差块依次增加到第 1 层的 5 个残差块。

第 1 层中，$F_{1\text{-}B}$ 通过 2 层卷积调整通道数量后映射为概率分布图再求最大值索引即得网络分割预测。

1.3　马尔可夫判别器

DbFFDA 的判别器为马尔可夫判别器[24]。如图 4 所示，分割预测输入判别器进行 4 次下采样后再进行 1 次卷积操作，即可输出通道数为 1 的特征图。马尔可夫判别器也称为 PatchGAN，输出特征图上的一点表征了分割预测上一块区域的真实性。判别器的优化方向是鉴别出输入判别器的分割预测属于源域预测还是目标域预测。分割网络的优化方向是生成判别器无法分辨的分割预测。两者相互博弈，不断提升自身性能，以达到对齐两域数据分布的目的。

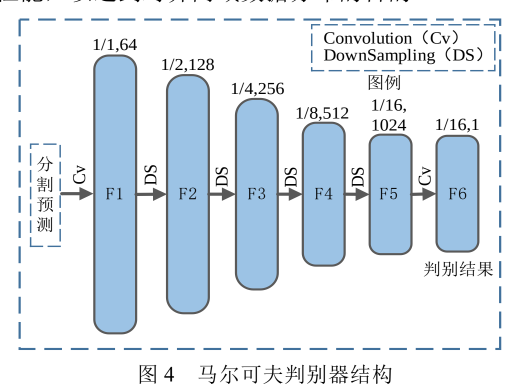


图 4　马尔可夫判别器结构

Fig. 4　Markov discriminator structure

1.4　融合特征输入

在复杂多变的自然光线条件下，受打印机打印质量及图像采集设备性能限制，表面压印字符图像具有明暗不均、噪点较多的特点。因此直接将原图像送入网络训练会给分割图像带来大量二值噪点，严重影响分割图像的纯净度。

将摄像机采集到的表面压印字符彩色图像转换为灰度图像作为网络输入数据的第 1 通道。

选用中值滤波的方法去除原图中的噪点，前景信息也最大程度地得到了保留。将中值滤波图像作为网络输入数据的第 2 通道。

表面压印字符图像中字符的边缘是前景的关键信息，对字符的分割具有重要意义。使用索贝尔(Sobel)算子对中值滤波图像进行边缘提取，所得图像作为网络输入数据的第 3 通道。

图 5 为网络输入的 3 通道数据示例图，将 3 通道图像拼接后送入网络进行训练。

	源域	目标域
通道 1：灰度图像	SDK0123	0A 04768
通道 2：中值滤波	SDK0123	0A 04768
通道 3：Sobel算子	SDK0123	0A 04768

图 5　融合特征输入

Fig. 5　Fusion feature input

1.5　分割连续性损失函数 L_{Con}

因表面压印字符图像噪点较多，部分字符扭曲变形，故仅使用交叉熵函数作为非对抗性损失函数对分割网络进行训练，目标域分割图像上出现了字符空洞与二值噪点。

为解决该问题，受传统计算机图像处理算法启发，提出分割连续性损失函数 L_{Con}。二值标签图中单个字符内部是连续的，背景亦连续，无二值噪点相互穿插，即前景与背景相对独立。因此 N 通道标签中各通道上值为 1 的点与值为 0 的点均应当是连续的，即某类与其他类相互独立且内部连续。如图 6 所示，除边界点外，某点与其邻域的点的值应当相同。

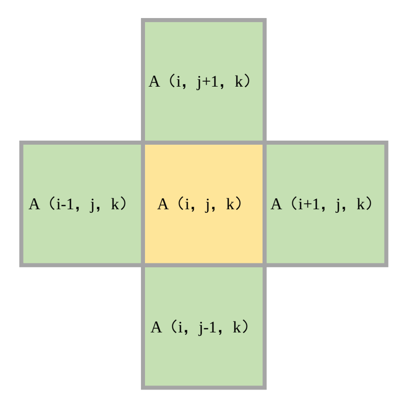


图 6　N 通道标签某通道中的某点及其四邻域

Fig. 6　Point of a channel in an n-channel label and its four neigh-borhoods

因网络分割图像由 N 通道特征图映射为概率分布图再求最大值索引得到，故网络输出的 N 通道特征图也应当具有与标签相同的性质，即各点的激活值连续。N 通道特征图某通道上的某点 $A_{(i,j,k)}$ 应具有与其邻域相近的激活值，故可构造分割连续性损失函数 L_{Con}：

$$L_{Con}=\frac{1}{N}\sum_{k=1}^{c}\sum_{j=1}^{m-1}\sum_{i=1}^{n-1}\begin{Bmatrix}\left|A_{(i,j,k)}-A_{(i-1,j,k)}\right|+\left|A_{(i,j,k)}-A_{(i+1,j,k)}\right|\\+\left|A_{(i,j,k)}-A_{(i,j-1,k)}\right|+\left|A_{(i,j,k)}-A_{(i,j+1,k)}\right|\end{Bmatrix} \tag{4}$$

其中，$A_{(i,j,k)}$ 代表第 k 通道上位于 (i,j) 处的点，$A_{(i-1,j,k)}$、$A_{(i+1,j,k)}$、$A_{(i,j-1,k)}$、$A_{(i,j+1,k)}$ 分别为该点的左、右、上、下邻域。该函数的引入有效抑制了字符中空洞的产生。

2　实验验证

2.1　实验数据

域适应分割有效解决了语义分割数据集手工标注成本高昂的问题，具有广阔的应用前景。在工业应用中，众多应用场景需要使用语义分割算法对采集得到的图像进行分割以进行后续处理。

在石墨电极生产这一工业场景中，为方便对生产过程中的石墨电极进行统计与管理，需要对石墨电极表面压印字符进行识别。语义分割是识别前的重要步骤，而语义分割数据集像素级的标注将会给企业带来巨大的生产成本。本文使用计算机生成源域数据，摄像机采集目标域数据，采用域适应分割方法有效地解决了这一问题，为企业节省了标注成本。

实验数据集包含源域数据和目标域数据两部分。源域(模拟)数据由计算机字库字符截图构成，无须手工标注，具有计算机生成的完备的标签集。图像中字符的字体、大小和空间位置均是随机的。目标域数据由手机摄像头采集某碳材料生产企业的石墨电极表面压印字符图像构成，如图 7 所示。

实验数据集示例图像如图 8 所示。源域数据(计算机字库字符)包含 600 张图像，目标域数据(石墨电极表面压印字符)包含 550 张图像，其中 440 张图像为训练数据，110 张图像为测试数据。数据集所有图像的尺寸均为 512×128。



图 7　石墨电极实拍图像

Fig. 7　Real image of graphite electrode

图 8　数据集图像示例

Fig. 8　Examples of dataset

2.2　评价指标

使用像素精度(Pixel Accuracy，PA)、平均像素精度(Mean Pixel Accuracy，MPA)、精准率(Precision)、召回率(Recall)和平均交并比(Mean Intersection Over Union，MIoU)等五种评价指标对网络性能进行评估。

PA 为分类正确的像素点数与总像素点数之比。*MPA* 为各类目标分类正确的像素点数与该类总像素点数之比的均值。*Precision* 为预测为正例的像素点中实际为正例的比例。*Recall* 为正确预测的正例的比例。交并比 *IoU* 为预测图像与标签图像中某一类物体交集所含像素点数与并集所含像素点数之比，平均交并比 *MIoU* 指各类交并比的均值。五种评价指标公式如下：

$$PA=\frac{TP+TN}{TP+FN+TN+FP} \tag{5}$$

$$MPA=mean\left(\frac{TP+TN}{TP+FN+TN+FP}\right) \tag{6}$$

$$Precision=\frac{TP}{TP+FP} \tag{7}$$

$$Recall=\frac{TP}{TP+FN} \tag{8}$$

$$MIoU=mean\left(\frac{A\cap B}{A\cup B}\right) \tag{9}$$

其中，真正例 *TP* 为实际为正，预测为正的像素点。真负例 *TN* 为实际为负，预测为负的像素点。假正例 *FP* 为实际为负，预测为正的像素点。假负例 *FN* 为实际为正，预测为负的像素点。

2.3　实验参数设置

实验依托 TensorFlow 平台构建神经网络，版本为 1.15，Python 版本为 3.6.5。实验平台配置如下：CPU 为 Intel E7-4830 v4，内存容量为 48G，GPU 为 Tesla K40c。

训练策略如下：在每一轮次训练中，从源域 600 张计算机字库字符图像中随机抽取 440 张图像及对应标签与目标域训练集 440 张石墨电极表面压印字符图像构成当前轮次训练数据。训练过程中对于每 1 个批次的源域数据，使用随机放缩及加噪的方法增加样本多样性，提升分割效果。生成器学习率设为 1e-5，判别器学习率设为 4e-6，使用 Adam 优化器进行优化。每迭代 1 轮保存 1 次网络模型，共迭代 100 轮。使用保存的网络模型对测试集进行测试，记录各轮模型测试结果。

2.4　实验结果与分析

2.4.1 对比实验结果与分析

实验分为 CLAN[10]、pOSAL[11]、BEAL[12]、DbFFDA 等 4 组。对比了 DbFFDA 与现有算法的性能表现。

MIoU 是语义分割领域使用最广泛的性能评价指标。选用 MIoU 作为主评价指标，找出各实验组的 MIoU 最大值对应训练轮次，将该轮评价指标数据作为实验结果。分割效果对比如图 9 所示，分割细节对比如图 10 所示。DbFFDA 与现有算法定量评价对比如表 1。

根据图 9、10 和表 1 数据可以看出：

a) CLAN 组分割所得图像中字符缺失严重，且存在大量块状噪点。无法进行后续识别，MIoU 仅为 56.22%。

b) pOSAL 组分割所得图像中字符较为完整，但仍存在大量块状噪点，MIoU 可达 66.83%。

c) BEAL 组块状噪点较 pOSAL 组具有一定程度的减少，但字符仍存在一定程度的缺失，MIoU 可达 68.46%。

d) DbFFDA 组分割所得图像中字符基本无缺失，边缘光滑无毛刺，块状噪点最少。MIoU 可达 69.60%，基本满足工业应用中表面压印字符后续识别需求。

图 9　分割效果对比

Fig. 9　Comparison of segmentation effect

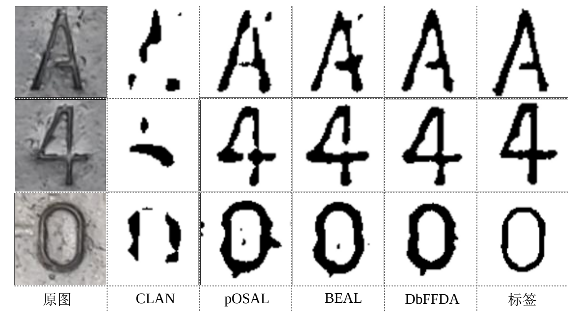


图 10　分割细节对比

Fig. 10　Segmentation detail comparison chart

表 1　对比实验定量评价

Tab. 1　Quantitative evaluation of comparative experiments　/%

Approachs	PA	MPA	Precision	Recall	MIoU
CLAN[10]	93.99	60.44	45.81	22.49	56.22
pOSAL[11]	93.51	**82.39**	45.43	64.14	66.83
BEAL[12]	95.59	82.07	46.91	61.34	68.46
DbFFDA	**96.12**	80.90	**59.60**	**64.29**	**69.60**

2.4.2 消融实验结果与分析

消融实验分为 U-Net(全监督)[2]、U-Net(无监督)[2]、Res-Adp(创新点 1)、Res-Adp+融合特征输入(创新点 1+2)、Res-Adp+融合特征输入+ L_{Con} (DbFFDA，创新点 1+2+3)等 5 组。对比了

DbFFDA 与全监督 U-Net 网络的性能差距，并实验证实了提出的各创新点的效果。找出各实验组的 MIoU 最大值对应训练轮次，将该轮评价指标数据作为实验结果。分割效果对比如图 11 所示，分割细节对比如图 12 所示。消融实验定量评价对比如表 2。

图 11　消融实验分割效果对比

Fig. 11　Image of segmentation effect comparison of ablation experiment

图 12　消融实验分割细节对比

Fig. 12　Segmentation detail comparison of ablation experiments

表 2　消融实验定量评价

Tab. 2　Quantitative evaluation of ablation experiment　/%

Appraochs	PA	MPA	Precision	Recall	MIoU
U-Net(全监督)[2]	98.31	90.67	79.01	82.30	83.08
U-Net(无监督)[2]	96.38	60.44	54.71	56.79	67.48
Res-Adp	95.75	84.78	51.87	72.85	69.00
Res-Adp+融合特征输入	96.73	79.20	60.65	60.11	69.43
Res-Adp+融合特征输入+ L_{Con}	96.12	80.90	59.60	64.29	69.60

根据图 11、12 及表 2 数据可以看出：a) 无监督 U-Net 组分割所得图像存在字符缺失、噪点较多等问题。Res-Adp 组(创新点 1)由于采用了双分支上采样结构，在对齐两域特征的同时又较为完整的保留了本域特征，极大程度地改善了这一问题，分割得到的字符连续无缺失，MIoU 较 U-Net(无监督)组提升了 1.52%。b) Res-Adp+融合特征输入组(创新点 1+2)由于输入网络的数据中增加了中值滤波图像与 Sobel 算子边缘检测图像，抑制了原图中的噪点，增强了字符的边缘信息。分割所得字符边缘较为光滑，抑制了毛刺现象的产生，且噪点较 Res-Adp 组(创新点 1)也有一定程度的减少，使得 MIoU 较 Res-Adp 组提升了 0.43%。c) Res-Adp+融合特征输入+ L_{Con} 组(DbFFDA，创新点 1+2+3)分割所得图像中字符连续无缺失，边缘光滑无毛刺，且背景中噪点最少。由于分割连续性损失函数 L_{Con} 的引入，进一步抑制了背景中的噪点与字符中空洞的产生，分割效果最佳，MIoU 可达 69.60%。

DbFFDA 与全监督 U-Net 网络在客观评价指标方面仍存在一定差距，但实际分割效果已基本满足字符识别需要，有望投入实际的工业应用中。

3　结束语

针对工业应用中表面压印字符图像对应标签获取困难，手工标注成本高昂的问题，本文提出一种双支路特征融合的域适应分割方法 DbFFDA。首先，借鉴 U-Net 的跨层连接设计思路，提出一种双分支上采样分割网络 Res-Adp。同时，将灰度图像、中值滤波图像与边缘检测图像融合作为网络输入数据，抑制了原图中的噪点，增强了字符的边缘信息。此外，因标签中各类对象内部连续，故分割网络输出特征图应当具有相同的性质，本文据此提出分割连续性损失函数 L_{Con}，该函数通过约束源域分割图像的生成间接提升目标域图像的分割效果，进一步抑制了字符中空洞与背景噪点的产生。DbFFDA 分割所得图像中字符完整、边缘光滑、噪点较少，MIoU 可达 69.60%，基本满足工业应用中表面压印字符后续识别需求。

针对部分图像因光照不均引起的分割预测中噪点过多的问题，作者将在后续工作中尝试对判别器结构进行改进，使其更好的约束生成器的优化，抑制分割预测中噪点的产生，提升分割效果。

参考文献：

[1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Press, 2015: 3431-3440.

[2] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [J]. Medical Image Computing and Computer-Assisted Intervention, 2015: 1520-1528.

[3] Chen L C, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected crfs [C]// International Conference in Learning Representations, 2015.

[4] Chen L C, Papandreou G, Kokkinos I, *et al.* Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018, 40 (4): 834-848.

[5] Chen L C, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation [J]. arXiv preprint arXiv: 1706. 05587, 2017.

[6] Chen L C, Zhu Yukun, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the European Conference on Computer Vision (ECCV) . 2018: 801-818.

[7] Goodfellow I J, Pouget A J, Mirza M, *et al.* Generative adversarial networks [C]// Proceedings of Annual Conference on Neural Information Processing Systems. Montreal, Quebec, Canada: ACM. 2014. 2672–2680.

[8] Ganin Y, Lempitsky V S. Unsupervised domain adaptation by backpropagation [C]// Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR, 2015. 1180–1189.

[9] Pei Zhongyi, Cao Zhangjie, Long Mingsheng, *et al.* Multi-adversarial domain adaptation [C]// Proceedings of the 32nd Conference on Artificial Intelligence, New Orleans, Louisiana, USA: AAAI, 2018. 3934–3941.

[10] Luo Yawei, Zheng Liang, Guan Tao, *et al.* Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2507-2516.

[11] Wang Shujun, Yu Lequan, Yang Xin, *et al.* Patch-based output space adversarial learning for joint optic disc and cup segmentation [J]. IEEE Trans on Medical Imaging, 2019, 38 (11): 2485-2495.

[12] Wang Shujun, Yu Lequan, Li Kang, *et al.* Boundary and entropy-driven adversarial learning for fundus image segmentation [C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 102-110.

[13] Zhu Junyan, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// 2017 IEEE International Conference on Computer Vision (ICCV) . 2017: 2242-2251.

[14] Chen Cheng, Dou Qi, Chen Hao, *et al.* Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation [J]. IEEE Trans on Medical Imaging, 2020, 39 (7): 2494-2505.

[15] 张勋晖，周勇，赵佳琦，等．基于熵增强的无监督域适应遥感图像语义分割 [J]．计算机应用研究，2021, 38 (09): 2852-2856. (Zhang Xunhui, Zhou Yong, Zhao Jiaqi, *et al.* Entropy enhanced unsupervised domain adaptive remote sensing image semantic segmentation [J]. Application Research of Computers, 2021, 38 (09): 2852-2856.)

[16] Liu Yuang, Zhang Wei, Wang Jun. Source-free domain adaptation for semantic segmentation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 1215-1224.

[17] Araslanov N, Roth S. Self-supervised augmentation consistency for adapting semantic segmentation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 15384-15394.

[18] Wang Qin, Dai Dengxin, Hoyer L, *et al.* Domain adaptive semantic segmentation with self-supervised depth estimation [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 8515-8525.

[19] Saha S, Obukhov A, Paudel D P, *et al.* Learning to relate depth and semantics for unsupervised domain adaptation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8197-8207.

[20] 刘少鹏，洪佳明，梁杰鹏，等．面向医学图像分割的半监督条件生成对抗网 [J]．软件学报, 2020, 31 (8): 2588−2602. (Liu Shaopeng, Hong Jiaming, Liang Jiepeng, *et al.* Medical image seg-mentation using semi-supervised conditional generative adversarial nets [J]. Journal of Software, 2020, 31 (8): 2588−2602.)

[21] Li Shuang, Xie Mixue, Gong Kaixiong, *et al.* Transferable semantic augmentation for domain adaptation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11516-11525.

[22] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proceedings of Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770–778.

[23] 胡扬涛，裴洋，林川，等．空洞残差 U 型网络用于视网膜血管分割 [J]．计算机工程与应用, 2021, 57 (07): 185-191. (Hu Yangtao, Pei Yang, Lin Chuan, *et al.* Atrous residual u-net for retinal vessel segmentation [J]. Computer Engineering and Applications, 2021, 57 (07): 185-191.)

[24] Isola P, Zhu Junyan, Zhou Tinghui, *et al.* Image-to-image translation with conditional adversarial networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.